The Gradient Projection Algorithm

1 Projections and Optimality Conditions

In this section we study the problem

$$\mathcal{P}: \quad \begin{array}{l} \min f(x) \\ \text{subject to } x \in \Omega \end{array}$$

where $\Omega \subset \mathbb{R}^n$ is assumed to be a nonempty closed convex set and f is C^1. The solution method that we will study is known as the gradient projection algorithm and was pioneered by Allen Goldstein of the University of Washington in 1964. In our theorem on first-order necessary conditions for optimality we found that if $\overline{x}$ is a local minimum for $\mathcal{P}$ then

$$\nabla f(\overline{x})^T(y - \overline{x}) \geq 0 \tag{1}$$

for all $y \in \Omega$. Moreover, if f is convex, then condition rfeqgp1 implies that $\overline{x}$ is a local minimum for $\mathcal{P}$. An instance of the function f that is of particular significance is

$$f(x) := \frac{1}{2}\|x - x_0\|_2^2.$$

In this case problem $\mathcal{P}$ becomes one of finding the closest point $\overline{x}$ in Ω to x_0. We now give the celebrated projection theorem for convex sets.

Theorem 1.1 *Let $x_0 \in \mathbb{R}^n$ and let $\Omega \subset \mathbb{R}^n$ be a nonempty closed convex set. Then $\overline{x} \in \Omega$ solves the problem*

$$\min\{\frac{1}{2}\|x - x_0\|_2^2 : x \in \Omega\}$$

if and only if

$$(\overline{x} - x_0)^T(y - \overline{x}) \geq 0 \tag{2}$$

for all $y \in \Omega$. Moreover, the solution $\overline{x}$ always exists and is unique.

Proof. Existence follows from the compactness of the set

$$\{x \in \Omega : \|x - x_0\|_2 \leq \|\hat{x} - x_0\|_2\}$$

where $\hat{x}$ is any element of Ω. Uniqueness follows from the strong convexity of the 2-norm squared. The remainder of the theorem follows immediately from Theorem **??** once it is observed that if

$$f(x) = \frac{1}{2}\|x - x_0\|_2^2$$

then

$$\nabla f(x) = x - x_0. \qquad \blacksquare$$

Definition 1.1 *Let $\Omega \subset \mathbb{R}^n$ be nonempty closed convex. We define the projection into Ω to be the mapping $P_\Omega : \mathbb{R}^n \to \Omega$ given by*

$$\frac{1}{2}\|P_\Omega(x) - x\|_2^2 = \min\{\frac{1}{2}\|y - x\|_2^2 : y \in \Omega\}.$$

Observe that P_Ω is well-defined by Theorem 1.1.

We now introduce two geometric concepts that aid in interpreting the optimality condition given in Theorem 1.1. Recall that the tangent cone to Ω at a point $x_0 \in \Omega$ is given by

$$T_\Omega(x_0) = \overline{\bigcup_{\lambda>0} \lambda(\Omega - x_0)}.$$

Dually, we call the set

$$N_\Omega(x) := \{z : \langle z, y - z\rangle \leq 0 \quad \text{for all } y \in \Omega\}$$

the normal cone to Ω at x.

Using the notions of a normal cone and a tangent cone we obtain the following restatement.

Theorem 1.2 *Let $\overline{x}$ be a solution to problem $\mathcal{P}$ and suppose that f is differentiable at $\overline{x}$, then*

$$-\nabla f(\overline{x}) \in N_\Omega(\overline{x}). \tag{3}$$

Moreover, if f is convex then (3) is sufficient for $\overline{x}$ to be a global minimizer of f on Ω.

Proof. We need only show that condition (3) is equivalent to the statement that

$$\nabla f(\overline{x})^T(y - \overline{x}) \geq 0 \quad \text{for all } y \in \Omega.$$

But this is clear from the definition of the normal cone. ∎

Theorem 1.3 *Let Ω be a non-empty closed convex subset of $\mathbb{R}^n$ and let P_Ω denote the projector into Ω. Then given $x \in \mathbb{R}^n$ we have $z = P_\Omega(x)$ if and only if*

$$(x - z) \in N_\Omega(z). \tag{4}$$

Proof. We need only show that (4) is equivalent to (2), but again this follows immediately from the definition of the normal cone. ∎

We have the following interesting corollary.

Corollary 1.3.1 *Let $x \in \Omega$, $z \in N_\Omega(x)$, and $t \geq 0$, then*

$$P_\Omega(x + tz) = x.$$

Proof. Simply observe that

$$(x + tz) - P_\Omega(x + tz) = tz \in N_\Omega(x),$$

so that the result follows from the theorem. ■

This yields the following corollary to Theorem 1.1 in the context of $\mathcal{P}$.

COROLLARY 1.3.2 *Let $\overline{x}$ be a solution to $\mathcal{P}$, then*

$$P_\Omega(\overline{x} - t\nabla f(\overline{x})) = \overline{x} \tag{5}$$

for all $t \geq 0$.

Proof. Just apply Theorem 1.1 and Corollary 1.3.1. ■

We now show how (5) can be used both as a stopping criteria for our algorithm and as a method for generating search directions.

PROPOSITION 1.2 *Let $x \in \Omega$ and set $d = P_\Omega(x - t\nabla f(x)) - x$. Then*

$$\nabla f(x)^T d \leq \frac{-\|P_\Omega(x - t\nabla f(x)) - x\|^2}{t}.$$

Proof. Simply observe that

$$\begin{aligned} &\|P_\Omega(x - t\nabla f(x)) - x\|^2 = \langle P_\Omega(x - t\nabla f(x)) - x, P_\Omega(x - t\nabla f(x)) - x\rangle \\ &= \quad -t\nabla f(x)^T d + \langle P_\Omega(x - t\nabla f(x)) - (x - t\nabla f(x)), P_\Omega(x - t\nabla f(x)) - x\rangle \\ &\leq \quad -t\nabla f(x)^T d \end{aligned}$$

where the last inequality follows Theorem 1.1 equation rfeqgp2. ■

Based on these observations we have the following algorithm.

2 The Basic Gradient Projection Method

Initialization: $x \in \Omega$, $\gamma \in (0, 1)$, $c \in (0, 1)$

Having x_k obtain x_{k+1} as follows

1. Set $d_k := P_\Omega(x_k - \nabla f(x_k)) - x_k$
2. Set

$$\begin{aligned} \lambda_k \quad := \quad & \max \gamma^s \\ & \text{subject to } s \in \{0, 1, 2, \ldots\} \\ & f(x_k) + \gamma^s d_k) - f(x_k) \leq c\gamma^s \nabla f(x_k)^T d_k. \end{aligned}$$

3. Set $x_{k+1} := x_k + \lambda_k d_k$.

We appeal to our result for the backtracking line search to yield a convergence theorem for this method.

THEOREM 2.1 *Let $f : \mathbb{R}^n \to R$ be C^1 and let $\Omega \subset \mathbb{R}^n$ be a nonempty closed convex set. Let $x_0 \in \Omega$ be such that f' is uniformly continuous on the set $\overline{co}\{x \in \Omega : f(x) \leq f(x_0)\}$. If $\{x_k\}$ is the sequence generated by gradient projection algorithm given above with starting point x_0, then one of the following must occur.*

1. *There is a k_0 such that $-\nabla f(x_{k_0}) \in N_\Omega(x_{k_0})$.*
2. *$f(x_k) \downarrow -\infty$.*
3. *The sequence $\{\|d_k\|\}$ diverges to $+\infty$,*
4. *For every subsequence $J \subset \mathbb{N}$ for which $\{d_k\}_J$ is bounded, we have that $d_k \underset{J}{\to} 0$, or equivalently*
$$\|P_\Omega(x_k - \nabla f(x_k)) - x_k\| \underset{J}{\to} 0.$$

COROLLARY 2.1.1 *Let the hypotheses of Theorem 2.1 hold. Furthermore assume that the sequence $\{d_k\}$ is bounded. Then every cluster point $\overline{x}$ of the sequence $\{x_k\}$ satisfies $-\nabla f(\overline{x}) \in N_\Omega(\overline{x})$.*

3 The Computation of Projections

We now address the question of implementation. Specifically, how does one compute the projection onto the convex set Ω. In general this is not a finite process. Nonetheless, for certain important convex sets Ω it can be done quite efficiently.

Projection onto box constraints

Let us suppose that Ω is given by $\Omega := \{x \in \mathbb{R}^n : \ell \leq x \leq u\}$, where $\ell, u \in \overline{\mathbb{R}}^n$ with $\overline{\mathbb{R}} = \Omega \cup \{+\infty, -\infty\}$ and $\ell :\leq u$, $i = 1, \ldots, n, \ell_i \neq +\infty$ $i = 1, \ldots, n$ and $u_i \neq -\infty$ $i = 1, \ldots, n$. Then P_Ω can be expressed componentwise as

$$[P_\Omega(x)]_i := \begin{cases} \ell_i & \text{if } x_i \leq \ell_i \\ x_i & \text{if } \ell_i < x_i < u_i \\ u_i & \text{if } u_i \leq x_i \end{cases}$$

Thus, for example, if $\Omega = \mathbb{R}^n_+$, then

$$P_\Omega(x) = x_+.$$

Projection onto a Polyhedron

Let Ω be the polyhedron given by

$$\Omega := \{x \in \mathbb{R}^n : a_i^T x \leq \alpha_i, i = 1, \ldots, 3, a_i^T x = \alpha_i, i = s+1, \ldots, m\}.$$

Then P_Ω is determined by solving the quadratic program

$$\begin{array}{lll} \min \frac{1}{2}\|x - y\|_2^2 & & \\ \text{subject to} & a_i^T x \leq \alpha_i & i = 1, \ldots, s \\ & a_i^T x = \alpha_i & i = s+1, \ldots, m. \end{array}$$